KELLOGG, HANSEN, TODD, FIGEL & FREDERICK, P.L.L.C.
SUMNER SQUARE
1615 M STREET, N.W.
SUITE 400
WASHINGTON, D.C. 20036-3215
(202) 326-7900
FACSIMILE:
(202) 326-7999

April 14, 2025

Judge William H. Alsup
Phillip Burton Federal Building & U.S. Courthouse
450 Golden Gate Avenue, 16th Floor Clerk's Office
San Francisco, CA 94102

Re: *X Corp. v. Bright Data Ltd.*, 3:23-cv-03698-WHA (N.D. Cal.)

Dear Judge Alsup:

I write for X in response to Bright Data's letter (Dkt. 244) seeking (1) to compel X to produce unnecessary platform data via an unworkable "data protocol"; and (2) a 30(b)(6) deposition that does not count against the deposition limit. Each request should be denied.[1] Bright Data rushed to file this motion so it could distract from its far-more-serious failure to produce its own scraping activity logs – logs it should have produced months ago and whose production X is moving to compel today. Indeed, Bright Data and its customers are the only ones that know exactly how, when, and to what extent they intruded into X's systems. Those details – not the ancillary-at-best data Bright Data's "protocol" seeks – are the key to whether this "data scraper impaired [X's] servers." Dkt. 246. Yet Bright Data is withholding all those details, using this unworkable "protocol" as its excuse. The Court should deny the motion.

1. ***X is Already Producing Data on Server Strain Caused by Scraping.*** Bright Data bases its motion on the false assertion that X has refused to provide information or data about server strain caused by unauthorized access and scraping. In fact, X has produced voluminous information on those topics, including the detailed "internal investigation" memo that underpinned the complaint's server-strain allegations, Dkt. 170-1, ¶¶ 83-92; the "Botox" technical analyses supporting the complaint's scraping estimates, *id.* ¶ 87; multiple documents laying out X's server architecture; incident reports documenting scraping-related server failures; and 215,846 electronic documents from five custodians, including engineers who work on X's servers. X is also in the process of performing targeted searches through its Atlassian/Jira database, which tracks server strain and outages and X's remedial response. X has already produced a significant amount of Jira data, including reports relating to scraping incidents. It is now completing its comprehensive search of the Jira database for more data about scraping-

[1] The demands for information about X's data systems and an extra deposition do not arise from any served discovery request or deposition notice. Bright Data did not serve an interrogatory to identify X's data systems, and this motion was the first X ever saw any 30(b)(6) topics.

related server strain and will produce the results in short order. This is on top of the hundreds of thousands of additional emails and Slacks X is now reviewing from seven new custodians, whose production will be substantially complete by agreement on June 3, 2025.

2. ***Bright Data's Proposed Data Protocol Is Unjustified.*** The real point of Bright Data's motion is not to obtain information from X; it is to invent a "protocol" Bright Data can hide behind in refusing to produce its own scraping activity logs. This whole tit-for-tat protocol is unprecedented – Bright Data cites no case (about scraping or anything else) adopting one – and the Court should reject it here. Simply put, Bright Data's scraping data and X's platform data are not remotely equivalent. Bright Data should produce its full scraping logs – not just a cherrypicked sample, as detailed in X's concurrently filed motion. Meanwhile, X is producing from its Jira database every ticketed incident of server harm related to scraping or unauthorized access. Comparing Bright Data's scraping log against this Jira data – plus the many other reports and emails X is producing – will illuminate how Bright Data's conduct caused server harm. Bright Data insists, however, that X go further and extract a panoply of irrelevant data about millions of individual Bright Data-associated IP addresses. Much of this data is not maintained in the ordinary course. The data that does exist would be prohibitively expensive to compile.

First, Bright Data seeks (at 2) "Platform Engagement Information" about Bright Data's and its customers' interactions with X's platform. In other words, Bright Data is seeking information *about what Bright Data and its customers did* on X's platform. That is information Bright Data – not X – should produce. After all, Bright Data (and its customers) are the ones who chose which X endpoints to target, evaded CAPTCHAs, bypassed rate limits, and hid their conduct from X. In other scraping cases, the defendant scrapers kept logs showing specific URL endpoints targeted, whether scraping efforts were successful or blocked, and scraping failure rates. The failure to preserve those logs resulted in sanctions. *See*, *e.g.*, *hiQ Labs, Inc. v. LinkedIn Corp.*, 639 F. Supp. 3d 944, 972-73, 978-80 (N.D. Cal. 2022). Here, the Court should not force X to tell Bright Data about its own behavior. Nor has Bright Data made any argument why the IP-by-IP-level data it seeks about its millions of IPs is necessary. To the extent specific scraping episodes harmed X's servers, that information will be captured in X's other documents.

In any event, X does not keep in the ordinary course much of the "Platform Engagement Information" Bright Data seeks. As X has told Bright Data, it maintains a data log from the Twitter Front End ("TFE"), which reflects every instance of access to X's platform (from scrapers and non-scrapers alike) from the world-wide-web. In the ordinary course, the TFE data log contains one month of data, but X is currently retaining data (at great expense) for the last 12 months. The data log is roughly 30 petabytes (30 million gigabytes) and costs approximately $500,000 per month to host. It contains data pertaining to users' IP addresses, time stamps, the endpoints and URLs visited, and whether users were authenticated. But it does not identify scraping, nor does it contain the other server-strain information Bright Data seeks.

The engineering effort and cost required to extract the limited engagement information Bright Data seeks from the TFE data log would be prohibitive. There is no user interface, which means running queries for the information Bright Data seeks would require X's engineers to design bespoke code and incur tens (if not hundreds) of thousands of dollars in additional hosting

costs. We understand that an individualized query for the millions of IPs used by Bright Data (even a limited sample would have at least hundreds of thousands of IPs) would rank among the largest and most complex data-query projects X has ever undertaken. That burden would be amplified because Bright Data has used blanket "Highly Confidential" designations to block anyone at X from seeing the IPs, so it is unclear how any query could be run at all.[2]

Second, Bright Data requests IP-specific "Server Information," including for every IP: the identity of the server that the scraper accessed, whether there was server failure, and the percentage of requests attributable to Bright Data IP addresses compared to other server requests. X does not believe it has this sort of granular information, as any one user who accesses a particular endpoint URL may touch scores (if not hundreds) of individual servers. Nor is X aware of any existing data documenting every individual request that every user made to every X server: retaining that sort of granular data at scale for a platform as large as X's would be essentially impossible. Nor is such data needed, as X is already producing information about its system architecture, plus server-strain data from its Jira system and other documents.

Third, Bright Data requests "User/Customer Information," consisting of the X accounts associated with Bright Data's customers or IP addresses. Again, this is information Bright Data claims to have through its "robust 'Know Your Customer'" process.[3] And to the extent Bright Data seeks account information to determine whether its customers sought to purchase data through X's "API" program, X has already produced that information. In any event, the engineering cost to query X's customer database and other files for the millions of IPs Bright Data has used is enormous. X estimates it would take months to generate the requested data.

More broadly, Bright Data cannot articulate why any of its requested data is necessary or proportionate in light of the trove of other directly relevant data X is producing. If the Court orders X to produce any additional data from the TFE data log, Bright Data should pay for the search and production costs. *See Lifetouch Nat'l Sch. Studios, Inc. v. Moss-Williams*, 2013 WL 11235928, at *3 (N.D. Cal. Oct. 15, 2013) ("Cost-shifting is also appropriate when the discovery sought imposes an undue burden or expense on the responding party.").

3. *X Does Not Oppose a Rule 30(b)(6) Deposition, but It Should Count Toward the Parties' Deposition Limits*. X has no objection to a deposition proceeding on a mutually agreed date, but any such deposition should count toward Bright Data's limit under the Court's standing order. If the Court awards an extra 30(b)(6) deposition on this topic, X also requests a 30(b)(6) deposition on Bright Data's more-serious data issues referenced in the attached Exhibit.

Respectfully submitted,
/s/ Joshua D. Branson
Counsel for X Corp.
**Admitted Pro Hac Vice*

[2] Bright Data also seeks data about interactions that millions of IP addresses had with X's platform *outside* of its scraping sessions. None of that information is remotely relevant.
[3] *See* https://perma.cc/WX2A-QD9C (describing Bright Data's customer diligence).